

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: HOFVANDER, Per
PERSSON, Per T
WIKSTROM, Olle
TALLBERG, Anneli

(ii) TITLE OF INVENTION: GENETICALLY ENGINEERED MODIFICATION OF
POTATO TO FORM AMYLOPECTIN-TYPE STARCH

(iii) NUMBER OF SEQUENCES: 21

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Burns, Doane, Swecker & Mathis
(B) STREET: George Mason Bldg., Washington & Prince Sts.
(C) CITY: Alexandria
(D) STATE: Virginia
(E) COUNTRY: United States
(F) ZIP: 22313-1404

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/070,455
(B) FILING DATE: 09-JUN-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Crane-Feury, Sharon E
(B) REGISTRATION NUMBER: 36,113
(C) REFERENCE/DOCKET NUMBER: 003300-293

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703) 836-6620
(B) TELEFAX: (703) 836-2021

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 342 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 217..342

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TGCATGTTTC CCTACATTCT ATTTAGAATC GTGTTGTGGT GTATAAACGT TGTTTCATAT	60
CTCATCTCAT CTATTCTGAT TTTGATTCTC TTGCCTACTG TAATCGGTGA TAAATGTGAA	120
TGCTTCCTTT CTTCTCAGAA ATCAATTCT GTTTTGT TTTT TGTTTCATCTG TAGCTTATTC	180

RECEIVED
JUN 02 1993
GROUP 1800

TCTGGTAGAT TCCCCTTTTT GTAGACCACA CATCAC ATG GCA AGC ATC ACA GCT	234
Met Ala Ser Ile Thr Ala	
1 5	
TCA CAC CAC TTT GTG TCA AGA AGC CAA ACT TCA CTA GAC ACC AAA TCA	282
Ser His His Phe Val Ser Arg Ser Gln Thr Ser Leu Asp Thr Lys Ser	
10 15 20	
ACC TTG TCA CAG ATA GGA CTC AGG AAC CAT ACT CTG ACT CAC AAT GGT	330
Thr Leu Ser Gln Ile Gly Leu Arg Asn His Thr Leu Thr His Asn Gly	
25 30 35	
TTA AGG GCT GTT	342
Leu Arg Ala Val	
40	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2549 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

B1
cont

AACAAGCTTG ATGGGCTCCA ATCAACAAC AATACTAAGG TAACACCCAA GATGGCATCC	60
AGAACTGAGA CCAAGAGACC TGGATGCTCA GCTACCATTG TTTGTGGAAA GGAATGAAC	120
TTGATCTTTG TGGGTACTGA GGTGGTCTCT TGGAGCAAAA CTGGTGGACT AGGTGATGTT	180
CTTGGTGGAC TACCACCAGC CCTTGCAGTA AGTCTTTCTT TCATTGTTGTT ACCTACTCAT	240
TCATTACTTA TTTTGTTTTAG TTAGTTTCTA CTGCATCAGT CTTTTTATCA TTTAGGCCCG	300
CGGACATCGG GTAATGACAA TATCCCCCGG TTATGACCAA TACAAAGATG CTTGGGATAC	360
TGGCGTTGCG GTTGAGGTAC ATCTTCCTAT ATTGATACGG TACAATATTG TTCTCTTACA	420
TTTCCTGATT CAAGAATGTG ATCATCTGCA GGTCAAAGTT GGAGACAGCA TTGAAATTGT	480
TCGTTTCTTT CACTGCTATA AACGTGGGGT TGATCGTGTT TTTGTTGACC ACCCAATGTT	540
CTTGGAGAAA GTAAGCATAT TATGATTATG AATCCGTCCT GAGGGATACG CAGAACAGGT	600
CATTTTGAGT ATCTTTTAAC TCTACTGGTG CTTTACTCT TTTAAGGTTT GGGGCAAAAC	660
TGGTTCAAAA ATCTATGGCC CCAAAGCTGG ACTAGATTAT CTGGACAATG AACTTAGGTT	720
CAGCTTGTTG TGTCAAGTAA GTTAGTTACT CTTGATTTTT ATGTGGCATT TTAATCTTTT	780
GTCTTTAATC GTTTTTTTTAA CCTGTGTTTC TCAGGCAGCC CTAGAGGCAC CTAAAGTTTT	840
GAATTTGAAC AGTAGCAACT ACTTCTCAGG ACCATATGGT AATTAACACA TCCTAGTTTC	900
AGAAAACCTC TTAATATATC ATTGTAGGTA ATCATCTTTA TTTTGCCTAT TCCTGCAGGA	960
GAGGATGTTT TCTTCATTGC CAATGATTGG CACACAGCTC TCATTCCTTG CTACTTGAAG	1020
TCAATGTACC AGTCCAGAGG AATCTACTTG AATGCCAAGG TAAATTTCT TTGTATTAC	1080
TCGATTGCAC GTTACCCTGC AAATCAGTAA GGTTGTATTA ATATATGATA AATTTACAT	1140

B1
cont

TGCCTCCAGG	TTGCTTTCTG	CATCCATAAC	ATTGCCTACC	AAGGTCGATT	TTCTTTCTCT	1200
GACTTCCCTC	TTCTCAATCT	TCCTGATGAA	TTCAGGGGTT	CTTTTGATTT	CATTGATGGG	1260
TATGTATTTA	TGCTTGAAAT	CAGACCTCCA	ACTTTTGAAG	CTCTTTTGAT	GCTAGTAAAT	1320
TGAGTTTTTA	AAATTTTGCA	GATATGAGAA	GCCTGTTAAG	GGTAGGAAAA	TCAACTGGAT	1380
GAAGGCTGGG	ATATTAGAAT	CACATAGGGT	GGTTACAGTG	AGCCCATACT	ATGCCCAAGA	1440
ACTTGCTCTCT	GCTGTTGACA	AGGGAGTTGA	ATTGGACAGT	GTCCTTCGTA	AGACTTGCAT	1500
AACTGGGATT	GTGAATGGCA	TGGATACACA	AGAGTGGAAC	CCAGCGACTG	ACAAATACAC	1560
AGATGTCAAA	TACGATATAA	CCACTGTAAG	ATAAGATTTT	TCCGACTCCA	GTATATACTA	1620
AATTATTTTG	TATGTTTATG	AAATTAAAGA	GTTCTTGCTA	ATCAAAATCT	CTATACAGGT	1680
CATGGACGCA	AAACCTTTAC	TAAAGGAGGC	TCTTCAAGCA	GCAGTTGGCT	TGCCTGTTGA	1740
CAAGAAGATC	CCTTTGATTG	GCTTCATCGG	CAGACTTGAG	GAGCAGAAAG	G TTCAGATAT	1800
TCTTGTTGCT	GCAATTCACA	AGTTCATCGG	ATTGGATGTT	CAAATGTAG	TCCTTGTAAG	1860
TACCAAATGG	ACTCATGGTA	TCTCTCTTGT	TGAGTTTACT	TGTGCCGAAA	CTGAAATTGA	1920
CCTGCTACTC	ATCCTATGCA	TCAGGGAAGT	GGCAAAAAGG	AGTTTGAGCA	GGAGATTGAA	1980
CAGCTCGAAG	TGTTGTACCC	TAACAAAGCT	AAAGGAGTGG	CAAAATTCAA	TGTCCCTTTG	2040
GCTCACATGA	TCACTGCTGG	TGCTGATTTT	ATGTTGGTTC	CAAGCAGATT	TGAACCTTGT	2100
GGTCTCATTC	AGTTACATGC	TATGCGATAT	GGAACAGTAA	GAACCAGAAG	AGCTTGTACC	2160
TTTTTACTGA	GTTTTTAAAA	AAAGAATCAT	AAGACCTTGT	TTCCATCTA	AAGTTTAATA	2220
ACCAACTAAA	TGTTACTGCA	GCAAGCTTTT	CATTTCTGAA	AATTGGTTAT	CTGATTTTAA	2280
CGTAATCACA	TGTGAGTCAG	GTACCAATCT	GTGCATCGAC	TGGTGGACTT	GTTGACACTG	2340
TGAAAGAAGG	CTATACTGGA	TTCCATATGG	GAGCCTTCAA	TGTTGAAGTA	TGTGATTTTA	2400
CATCAATTGT	GTA CTGTAC	ATGGTCCATT	CTCGTCTTGA	TATACCCCTT	GTTGCATAAA	2460
CATTA ACTTA	TTGCTTCTTG	AATTTGGTTA	GTGCGATGTT	GTTGACCCAG	CTGATGTGCT	2520
TAAGATAGTA	ACAACAGTTG	CTAGAGCTC				2549

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 492 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..15

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 101..218

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAG CTC TCC TGG AAG GTAAGTGTGA ATTTGATAAT TTGCGTAGGT ACTTCAGTTT	55
Glu Leu Ser Trp Lys	
1 5	
GTGTTCTCG TCAGCACTGA TGGATTCCAA CTGGTGTCT TGCAG GAA CCT GCC	109
Glu Pro Ala	
1	
AAG AAA TGG GAG ACA TTG CTA TTG GGC TTA GGA GCT TCT GGC AGT GAA	157
Lys Lys Trp Glu Thr Leu Leu Leu Gly Leu Gly Ala Ser Gly Ser Glu	
5 10 15	
CCC GGT GTT GAA GGG GAA GAA ATC GCT CCA CTT GCC AAG GAA AAT GTA	205
Pro Gly Val Glu Gly Glu Glu Ile Ala Pro Leu Ala Lys Glu Asn Val	
20 25 30 35	
GCC ACT CCT TAAATGAGCT TTGGTTATCC TTGTTTCAAC AATAAGATCA	254
Ala Thr Pro *	
TTAAGCAAAC GTATTTACTA GCGAACTATG TAGAACCCTA TTATGGGGTC TCAATCATCT	314
ACAAAATGAT TGGTTTTTGC TGGGGAGCAG CAGCATATAA GGCTGTAAAA TCCTGGTTAA	374
TGTTTTTGTA GGTAAGGGCT ATTTAAGGTG GTGTGGATCA AAGTCAATAG AAAATAGTTA	434
TTACTAACGT TTGCAACTAA ATACTTAGTA ATGTAGCATA AATAATACTA GAACTAGT	492

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 987 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGCTTTAAC GAGATAGAAA ATTATGTTAC TCCGTTTTGT TCATTACTTA ACAAATGCAA	60
CAGTATCTTG TACCAAATCC TTTCTCTCTT TTCAAACCTT TCTATTTGGC TGTTGACGGA	120
GTAATCAGGA TACAAACCAC AAGTATTTAA TTGACTCCTC CGCCAGATAT TATGATTTAT	180
GAATCCTCGA AAAGCCTATC CATTAAGTCC TCATCTATGG ATATACTTGA CAGTATCTTC	240
CTGTTTGGGT ATTTTTTTTT CCTGCCAAGT GGAACGGAGA CATGTTATGA TGTATACGGG	300
AAGCTCGTTA AAAAAAATA CAATAGGAAG AAATGTAACA AACATTGAAT GTTGTTTTTA	360
ACCATCCTTC CTTTAGCAGT GTATCAATTT TGTAATAGAA CCATGCATCT CAATCTTAAT	420
ACTAAATGC AACTTAATAT AGGCTAAACC AAGATAAAGT AATGTATTCA ACCTTTAGAA	480
TTGTGCATTC ATAATTAGAT CTTGTTTGTC GTAAAAAATT AGAAAATATA TTTACAGTAA	540
TTTGAATAC AAAGCTAAGG GGAAGTAAC TAATATTCTA GTGGAGGGAG GGACCAGTAC	600
CAGTACCTAG ATATTATTTT TAATTACTAT AATAATAATT TAATTAACAC GAGACATAGG	660
AATGTCAAGT GGTAGCGTAG GAGGGAGTTG GTTTAGTTTT TTAGATACTA GGAGACAGAA	720

CCGGACGGCC	CATTGCAAGG	CCAAGTTGAA	GTCCAGCCGT	GAATCAACAA	AGAGAGGGCC	780
CATAATACTG	TCGATGAGCA	TTTCCCTATA	ATACAGTGTC	CACAGTTGCC	TTCTGCTAAG	840
GGATAGCCAC	CCGCTATTCT	CTTGACACGT	GTCAGTAAA	CCTGCTACAA	ATAAGGCAGG	900
CACCTCCTCA	TTCTCACTCA	CTCACTCACA	CAGCTCAACA	AGTGGTAACT	TTTACTCATC	960
TCCTCCAATT	ATTTCTGATT	TCATGCA				987

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4964 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

B1
cont

AAGCTTTAAC	GAGATAGAAA	ATTATGTTAC	TCCGTTTGT	TCATTACTTA	ACAAATGCAA	60
CAGTATCTTG	TACCAAATCC	TTTCTCTCTT	TTCAAACCTT	TCTATTGGC	TGTTGACGGA	120
GTAATCAGGA	TACAAACCAC	AAGTATTTAA	TTGACTCCTC	CGCCAGATAT	TATGATTTAT	180
GAATCCTCGA	AAAGCCTATC	CATTAAGTCC	TCATCTATGG	ATATACTTGA	CAGTATCTTC	240
CTGTTTGGGT	ATTTTTTTTT	CCTGCCAAGT	GGAACGGAGA	CATGTTATGA	TGTATACGGG	300
AAGCTCGTTA	AAAAAAATA	CAATAGGAAG	AAATGTAACA	AACATTGAAT	GTTGTTTTTA	360
ACCATCCTTC	CTTTAGCAGT	GTATCAATTT	TGTAATAGAA	CCATGCATCT	CAATCTTAAT	420
ACTAAAATGC	AACTTAATAT	AGGCTAAACC	AAGATAAAGT	AATGTATTCA	ACCTTTAGAA	480
TTGTGCATTC	ATAATTAGAT	CTTGTTTGTC	GTAAAAAATT	AGAAAATATA	TTTACAGTAA	540
TTTGGAATAC	AAAGCTAAGG	GGGAAGTAAC	TAATATTCTA	GTGGAGGGAG	GGACCAGTAC	600
CAGTACCTAG	ATATTATTTT	TAATTACTAT	AATAATAATT	TAATTAACAC	GAGACATAGG	660
AATGTCAAGT	GGTAGCGTAG	GAGGGAGTTG	GTTTAGTTTT	TTAGATACTA	GGAGACAGAA	720
CCGGACGGCC	CATTGCAAGG	CCAAGTTGAA	GTCCAGCCGT	GAATCAACAA	AGAGAGGGCC	780
CATAATACTG	TCGATGAGCA	TTTCCCTATA	ATACAGTGTC	CACAGTTGCC	TTCTGCTAAG	840
GGATAGCCAC	CCGCTATTCT	CTTGACACGT	GTCAGTAAA	CCTGCTACAA	ATAAGGCAGG	900
CACCTCCTCA	TTCTCACTCA	CTCACTCACA	CAGCTCAACA	AGTGGTAACT	TTTACTCATC	960
TCCTCCAATT	ATTTCTGATT	TCATGCATGT	TTCCCTACAT	TCTATTATGA	ATCGTGTGTG	1020
GGTGATAAAA	CGTTGTTTCA	TATCTCATCT	CATCTATTCT	GATTTTGATT	CTCTGCCTA	1080
CTGTAATCGG	TGATAAATGT	GAATGCTTCC	TTTCTTCTCA	GAAATCAATT	TCTGTTTTGT	1140
TTTTGTTCAT	CTGTAGCTTA	TTCTCTGGTA	GATTCCCTT	TTGTAGACC	ACACATCACA	1200
TGGCAAGCAT	CACAGCTTCA	CACCACTTTG	TGTCAAGAAG	CCAACTTCA	CTAGACACCA	1260
AATCAACCTT	GTACAGATA	GGACTCAGGA	ACCATACTCT	GAATCACAAT	GGTTTAAGGG	1320

CTGTTAACAA GCTTGATGGG CTCCAATCAA CAACTAATAC TAAGGTAACA CCCAAGATGG 1380
CATCCAGAAC TGAGACCAAG AGACCTGGAT GCTCAGCTAC CATTGTTTGT GGAAAGGGAA 1440
TGAAC TTGAT CTTTGTGGGT ACTGAGGTTG GTCCTTGGAG CAAAAGTGGT GGACTAGGTG 1500
ATGTTCTTGG TGGACTACCA CCAGCCCTTG CAGTAAGTCT TTCTTTCATT TGGTTACCTA 1560
CTCATTCACT ACTTATTTTG TTTAGTTAGT TTCTACTGCA TCAGTCTTTT TATCATTTAG 1620
GCCCCGCGAC AGCGGGTAAT GACAATATCC CCCCCTTATG ACCAATACAA AGATGCTTGG 1680
GATACTGGCG TTGCGGTTGA GGTACATCTT CCTATATTGA TACGGTACAA TATTGTTCTC 1740
TTACATTTCC TGATTCAAGA ATGTGATCAT CTGCAGGTCA AAGTTGGAGA CAGCATTGAA 1800
ATTGTTTCGT TCTTTCCTG CTATAAACGT GGGGTTGATC GTGTTTTTGT TGACCACCCA 1860
ATGTTCTTGG AGAAAGTAAG CATATTATGA TTATGAATCC GTCCTGAGGG ATACGCAGAA 1920
CAGGTCATTT TGAGTATCTT TTAAGTCTAC TGGTGCTTTT ACTCTTTTAA GGTTTGGGGC 1980
AAAAGTGGT CAAAATCTA TGGCCCCAAA GCTGGACTAG ATTATCTGGA CAATGAAGTT 2040
AGGTTCAAGT TGTGTGTGCA AGTAAGTTAG TTAAGTCTGA TTTTATGTG GCATTTTACT 2100
CTTTTGTCTT TAATCGTTTT TTAAGCTTG TTTTCTCAGG CAGCCCTAGA GGCACCTAAA 2160
GTTTTGAATT TGAACAGTAG CAACTACTTC TCAGGACCAT ATGGTAATTA ACACATCCTA 2220
GTTTCAGAAA ACTCCTTACT ATATCATTGT AGGTAATCAT CTTTATTTTG CCTATTCCTG 2280
CAGGAGAGGA TGTCTCTTTC ATTGCCAATG ATTGGCACAC AGCTCTCATT CTTGCTACT 2340
TGAAGTCAAT GTACCAGTCC AGAGGAATCT ACTTGAATGC CAAGGTAAAA TTTCTTTGTA 2400
TTCCTCGAT TGCACGTTAC CCTGCAAATC AGTAAGTTG TATTAATATA TGATAAATTT 2460
CACATTGCCT CCAGGTTGCT TTCTGCATCC ATAACATTGC CTACCAAGGT CGATTTTCTT 2520
TCTCTGACTT CCCTCTTCTC AATCTTCTG ATGAATTCAG GGGTTCTTTT GATTTTCATTG 2580
ATGGGTATGT ATTTATGCTT GAAATCAGAC CTCCAAGTTT TGAAGCTCTT TTGATGCTAG 2640
TAAATTGAGT TTTTAAATTT TTGCAGATAT GAGAAGCCTG TTAAGGGTAG GAAATCAAC 2700
TGGATGAAGG CTGGGATATT AGAATCACAT AGGGTGGTTA CAGTGAGCCC ATACTATGCC 2760
CAAGAACTTG TCTCTGCTGT TGACAAGGGA GTTGAATTGG ACAGTGTCTT TCGTAAGACT 2820
TGCATAACTG GGATTGTGAA TGGCATGGAT ACACAAGAGT GGAACCCAGC GACTGACAAA 2880
TACACAGATG TCAAATACGA TATAACCACT GTAAGATAAG ATTTTCCGA CTCCAGTATA 2940
TACTAAATTA TTTTGTATGT TTATGAAATT AAAGAGTTCT TGCTAATCAA AATCTCTATA 3000
CAGGTCATGG ACGCAAACCT TTTACTAAAG GAGGCTCTTC AAGCAGCAGT TGGCTTGCTT 3060
GTTGACAAGA AGATCCCTTT GATTGGCTTC ATCGGCAGAC TTGAGGAGCA GAAAGGTTCA 3120
GATATTCTTG TTGCTGCAAT TCACAAGTTC ATCGGATTGG ATGTTCAAAT TGAGTTCCTT 3180
GTAAGTACCA AATGGACTCA TGGTATCTCT CTTGTTGAGT TTAAGTGTGC CGAACTGAA 3240
ATTGACCTGC TACTCATCCT ATGCATCAGG GAACTGGCAA AAAGGATTTT GAGCAGGAGA 3300
TTGAACAGCT CGAAGTGTTG TACCCTAACA AAGCTAAAGG AGTGGCAAAA TTCAATGTCC 3360

B1
cont

CTTTGGCTCA CATGATCACT GCTGGTGCTG ATTTTATGTT GGTTCCAAGC AGATTGGAAC 3420
 CTTGTGGTCT CATTCAAGTTA CATGCTATGC GATATGGAAC AGTAAGAACC AGAAGAGCTT 3480
 GTACCTTTTT ACTGAGTTTT TAAAAAAGA ATCATAAGAC CTTGTTTTCC ATCTAAAGTT 3540
 TAATAACCAA CTAAATGTGA CTGCAGCAAG CTTTTCATTT CTGAAAATTG GTTATCTGAT 3600
 TTTAACGTAA TCACATGTGA GTCAGGTACC AATCTGTGCA TCGACTGGTG GACTTGTTGA 3660
 CACTGTGAAA GAAGGCTATA CTGGATTCCA TATGGGAGCC TTCAATGTTG AAGTATGTGA 3720
 TTTTACATCA ATTGTGTACT TGTACATGGT CCATTCTCGT CTTGATATAC CCCTTGTTGC 3780
 ATAAACATTA ACTTATTGCT TCTTGAATTT GGTTAGTGCG ATGTTGTTGA CCCAGCTGAT 3840
 GTGCTTAAGA TAGTAACAAC AGTTGCTAGA GCTCTTGCG TCTATGGCAC CCTCGCATTT 3900
 GCTGAGATGA TAAAAAATTG CATGTCAGAG GAGCTCTCCT GGAAGGTAAG TGTGAATTTG 3960
 ATAATTTGCG TAGGTACTTC AGTTTGTGTT TCTCGTCAGC ACTGATGGAT TCCAAGTGGT 4020
 GTTCTTGCGAG GAACCTGCCA AGAAATGGGA GACATTGCTA TTGGGCTTAG GAGCTTCTGG 4080
 CAGTGAACCC GGTGTTGAAG GGGAAGAAAT CGCTCCACTT GCCAAGGAAA ATGTAGCCAC 4140
 TCCTTAAATG AGCTTTGGTT ATCCTTGTTT CAACAATAAG ATCATTAAGC AAACGTATTT 4200
 ACTAGCGAAC TATGTAGAAC CCTATTATGG GGTCTCAATC ATCTACAAA TGATTGGTTT 4260
 TTGCTGGGGA GCAGCAGCAT ATAAGGCTGT AAAATCCTGG TTAATGTTTT TGTAGGTAAG 4320
 GGCTATTTAA GGTGGTGTGG ATCAAAGTCA ATAGAAAATA GTTATTACTA ACGTTTGCAA 4380
 CTAAATACTT AGTAATGTAG CATAAATAAT ACTAGAACTA GTAGCTAATA TATATGCGTG 4440
 AATTTGTTGT ACCTTTTCTT GCATAATTAT TTGCAGTACA TATATAATGA AAATTACCCA 4500
 AGGAATCAAT GTTCTTGCT CCGTCCTCCT TTGATGATTT TTTACGCAAT ACAGAGCTAG 4560
 TGTGTTATGT TATAAATTTT GTTTAAAAGA AGTAATCAAA TTCAAATTAG TTGTTTGGTC 4620
 ATATGAAAGA AGCTGCCAGG CTAACCTTGA GGAGATGGCT ATTGAATTTT AAAATGATTA 4680
 TGTGAAAACA ATGCAACATC TATGTCAATC AACACTTAAA TTATTGCATT TAGAAAGATA 4740
 TTTTGTAGCC CATGACACAT TCATTCATAA AGTAAGGTAG TATGTATGAT TGAATGGACT 4800
 ACAGCTCAAT CAAAGCATCT CCTTTACATA ACGGCACTGT CTCTTGTCTA CTAATCTATT 4860
 GGTAGTAGTA GTAGTAATTT TACAATCCAA ATTGAATAGT AATAAGATGC TCTCTATTTA 4920
 CTAAAGTAGT AGTATTATTC TTTCGTTACT CTAAAGCAAC AAAA 4964

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 69 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site

- (B) LOCATION: 1..69
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1-207 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Lys Leu Asp Gly Leu Gln Ser Thr Thr Asn Thr Lys Val Thr Pro
1 5 10 15
Lys Met Ala Ser Arg Thr Glu Thr Lys Arg Pro Gly Cys Ser Ala Thr
20 25 30
Ile Val Cys Gly Lys Gly Met Asn Leu Ile Phe Val Gly Thr Glu Val
35 40 45
Gly Pro Trp Ser Lys Thr Gly Gly Leu Gly Asp Val Leu Gly Gly Leu
50 55 60
Pro Pro Ala Leu Ala
65

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
(B) LOCATION: 1..27
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 296-377 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Ala Arg Gly His Arg Val Met Thr Ile Ser Pro Arg Tyr Asp Gln Tyr
1 5 10 15
Lys Asp Ala Trp Asp Thr Gly Val Ala Val Glu
20 25

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
(B) LOCATION: 1..33
(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 452-550 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Lys Val Gly Asp Ser Il Glu Ile Val Arg Phe Phe His Cys Tyr
1 5 10 15
Lys Arg Gly Val Asp Arg Val Phe Val Asp His Pro Met Phe Leu Glu
20 25 30
Lys

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..30
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 647-736 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Val Trp Gly Lys Thr Gly Ser Lys Ile Tyr Gly Pro Lys Ala Gly Leu
1 5 10 15
Asp Tyr Leu Asp Asn Glu Leu Arg Phe Ser Leu Leu Cys Gln
20 25 30

B1
cont
(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..21
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 815-878 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ala Ala Leu Glu Ala Pro Lys Val Leu Asn Leu Asn Ser Ser Asn Tyr
1 5 10 15
Phe Ser Gly Pro Tyr
20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1..34

(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 878 and 959-1059 of SEQ ID NO. 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Gly Glu Asp Val Leu Phe Ile Ala Asn Asp Trp His Thr Ala Leu Ile
1 5 10 15

Pro Cys Tyr Leu Lys Ser Met Tyr Gln Ser Arg Gly Ile Tyr Leu Asn
20 25 30

Ala Lys

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 38 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1..38

(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1150-1263 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Val Ala Phe Cys Ile His Asn Ile Ala Tyr Gln Gly Arg Phe Ser Phe
1 5 10 15

Ser Asp Phe Pro Leu Leu Asn Leu Pro Asp Glu Phe Arg Gly Ser Phe
20 25 30

Asp Phe Ile Asp Gly Tyr
35

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1..79

(D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1349-1585 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

Lys Pro Val Lys Gly Arg Lys Ile Asn Trp Met Lys Ala Gly Ile Leu
1           5           10           15
Glu Ser His Arg Val Val Thr Val Ser Pro Tyr Tyr Ala Gln Glu Leu
20           25           30
Val Ser Ala Val Asp Lys Gly Val Glu Leu Asp Ser Val Leu Arg Lys
35           40           45
Thr Cys Ile Thr Gly Ile Val Asn Gly Met Asp Thr Gln Glu Trp Asn
50           55           60
Pro Ala Thr Asp Lys Tyr Thr Asp Val Lys Tyr Asp Ile Thr Thr
65           70           75

```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..59
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1676-1855 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```

Val Met Asp Ala Lys Pro Leu Leu Lys Glu Ala Leu Gln Ala Ala Val
1           5           10           15
Gly Leu Pro Val Asp Lys Lys Ile Pro Leu Ile Gly Phe Ile Gly Arg
20           25           30
Leu Glu Glu Gln Lys Gly Ser Asp Ile Leu Ala Val Ala Ile His Lys
35           40           45
Phe Ile Gly Leu Asp Val Gln Ile Val Val Leu
50           55

```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..64
- (D) OTHER INFORMATION: /note= "Amin acid sequence encoded by nucleotides 1945-2136 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gly Thr Gly Lys Lys Glu Phe Glu Gln Glu Ile Glu Gln Leu Glu Val
 1 5 10 15
 Leu Tyr Pro Asn Lys Ala Lys Gly Val Ala Lys Phe Asn Val Pro Leu
 20 25 30
 Ala His Met Ile Thr Ala Gly Ala Asp Phe Met Leu Val Pro Ser Arg
 35 40 45
 Phe Glu Pro Cys Gly Leu Ile Gln Leu His Ala Met Arg Tyr Gly Thr
 50 55 60

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..29
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 2301-2386 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Val Pro Ile Cys Ala Ser Thr Gly Gly Leu Val Asp Thr Val Lys Glu
 1 5 10 15
 Gly Tyr Thr Gly Phe His Met Gly Ala Phe Asn Val Glu
 20 25

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..19
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 2492-2459 of SEQ ID NO 2."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys Asp Val Val Asp Pro Ala Asp Val Leu Lys Ile Val Thr Thr Val
 1 5 10 15
 Ala Arg Ala

B1
cont

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 111 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..111
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 1200-1532 of SEQ ID NO 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```

Met Ala Ser Ile Thr Ala Ser His His Phe Val Ser Arg Ser Gln Thr
1           5           10           15
Ser Leu Asp Thr Lys Ser Thr Leu Ser Gln Ile Gly Leu Arg Asn His
20          25          30
Thr Leu Thr His Asn Gly Leu Arg Ala Val Asn Lys Leu Asp Gly Leu
35          40          45
Gln Ser Thr Thr Asn Thr Lys Val Thr Pro Lys Met Ala Ser Arg Thr
50          55          60
Glu Thr Lys Arg Pro Gly Cys Ser Ala Thr Ile Val Cys Gly Lys Gly
65          70          75          80
Met Asn Leu Ile Phe Val Gly Thr Glu Val Gly Pro Trp Ser Lys Thr
85          90          95
Gly Gly Leu Gly Asp Val Leu Gly Gly Leu Pro Pro Ala Leu Ala
100         105         110

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 43 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..43
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 3817-3945 of SEQ ID NO. 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

Cys Asp Val Val Asp Pro Ala Asp Val Leu Lys Ile Val Thr Thr Val
1           5           10           15
Ala Arg Ala Leu Ala Val Tyr Gly Thr Leu Ala Phe Ala Glu Met Ile
20          25          30
Lys Asn Cys Met Ser Glu Glu Leu Ser Trp Lys
35          40

```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amin acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 1..38
- (D) OTHER INFORMATION: /note= "Amino acid sequence encoded by nucleotides 4031-4144 of SEQ ID NO. 5."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Glu Pro Ala Lys Lys Trp Glu Thr Leu Leu Leu Gly Leu Gly Ala Ser
1 5 10 15

Gly Ser Glu Pro Gly Val Glu Gly Glu Ile Ala Pro Leu Ala Lys
20 25 30

Glu Asn Val Ala Thr Pro
35

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(ix) FEATURE:

- (A) NAME/KEY: misc_RNA
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /note= "Nucleotide 1 is a 7-methyl guanine added by 5'-5' linkage as an RNA cap."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAUGGCAAGA AAAAAAA